

DAFTAR PUSTAKA

- Abraham, R., 2017. Pemilihan Bahasa dan Presentasi Diri dalam Media Sosial Instagram di Kalangan Mahasiswa Program Studi Inggris Universitas Indonesia. *Universitas Indonesia*, pp. 1-18.
- Boughorbel, S., Jarray, F. & El-Anbari, M., 2017. Optimal classifier for imbalanced data using Matthew Correlation Coefficient metric. *PLoS ONE*, pp. 1-17.
- Chung-Hong Lee, H.-C. Y., 2009. Construction of Supervised and Unsupervised Learning Systems for Multilingual Text Categorization. *Expert Systems with Applications*, 36(2), pp. 2400-2410.
- Conneau, A. & Khandelwal, K., 2020. Unsupervised Cross-lingual Representation Learning at Scale. *Facebook AI*, pp. 1-12.
- Daskalaki, S., Kopanas, I. & Avouris, N., 2006. Evaluation of Classifier for an Uneven Class Distribution Problem. *Applied Artificial Intelligence*, 20(5), pp. 381-417.
- Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformer for Language Understanding. *Google AI Language*, pp. 1-16.
- Eberhard, M., D., Simon, G. F. & Fennig, C. D., 2021. *Ethnologue: Languages of the World*. Twenty-fourth edition ed. Dallas, Texas: SIL International.
- Gonalves, T. & Quaresma, P., 2010. Multilingual Text Classification through Combination of Monolingual Classifiers. *Workshop on Legal Ontologies and Artificial Intelligence Techniques*, pp. 29-38.
- Kemp, S., 2020. *Digital 2020: Indonesia*. [Online] Available at: <https://datareportal.com/reports/digital-2020-indonesia>
- Kudo, T. & Richardson, J., 2018. SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. *EMNLP*, pp. 1-6.
- Lample, G. & Conneau, A., 2019. Cross-lingual Language Model Pretraining. *Facebook AI*, pp. 1-10.
- Leonardo Rigutini, M. M. B. L., 2005. An EM Based Training Algorithm for Cross-Language Text Categorization. *The 2005 IEEE/WIC/ACM/ International Conference on Web Intelligence (WI'05)*, pp. 529-535.
- Liddy & D., E., 2001. Natural Language Processing. *In Encyclopedia of Library and Information Science*.

Liu, Y. et al., 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *Facebook AI*, pp. 1-13.

M. Ikonomakis, S. K. V. T., 2005. Text Classification Using Machine Learning Techniques. *WSEAS TRANSACTIONS on COMPUTERS*, p. 1.

Magueresse, A., Carles, V. & Heetderks, E., 2012. Low-resource Languages: A Review of Past Work and Future Challenges. *Department of Computer Science and Technology Tsinghua University*, p. 1.

Murakami, Y., 2019. Indonesia Language Sphere: an ecosystem for dictionary development for low-resource languages. *Journal of Physics: Conference Series*, pp. 1-8.

Nuria Bel, C. H. K. M. V., 2003. Cross-Lingual Text Categorization. *Research and Advanced Technology for Digital Libraries*, pp. 126-139.

Willie, B. et al., 2020. IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*, pp. 843-857.